

A method for spatializing audio in conference calls is disclosed in which the participants (A-E) in the calls are placed in particular locations in order to provide an additional dimension (direction) so that the participants (A-E) can better recognize who is speaking. Also, the dimension of volume is provided, which can be used in creating background sub-conferences (304-312). Consequently, sub-conferences (302-312) can be conducted in which the participants (A-E) can move seamlessly between them and also listen to other sub-conferences (304-312) being conducted in the background.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

**METHOD AND SYSTEM FOR
PROVIDING SPATIALIZED AUDIO
IN CONFERENCE CALLS**

5 BACKGROUND OF THE INVENTION

Technical Field of the Invention

The present invention relates generally to the telecommunications field and, in particular, to a method and system for providing spatialized audio in conference calls.

10 Description of Related Art

Conference calls are becoming an increasingly common communications medium. For example, a large corporation can have offices located throughout the world, but the corporation's employees at different locations are often required to consult with each other by conference call, in order to develop conclusions and solutions for pressing problems. Furthermore, the younger generation's current use of Internet chat rooms for "text-chats" will likely extend that practice to "voice-chats" (i.e., conference calls).

The conventional conference call systems in use today utilize a single voice channel for all participants, and a moderator typically controls the conference calls. As such, an individual can participate in only one conference call at a time.

A number of significant problems exist with the existing approaches taken for making conference calls. For example, during a conference call, it is often difficult to recognize who is speaking by voice alone. This recognition problem can be exacerbated if there are several participants in the conference call with similar regional accents or voices that sound similar. Furthermore, two or more conference call participants can be speaking at the same time, which degrades the conversations. Another problem with existing conference call approaches is that they do not make it possible to divide a conference call into a number of sub-conferences, and also to allow participants to move freely between the sub-conferences. Yet another problem with existing conference call approaches is that

-2-

a moderator needs to be appointed whenever the number of participants exceeds a certain limit. However, as described in detail below, the present invention successfully resolves the above-described problems.

SUMMARY OF THE INVENTION

5 In accordance with the present invention, a method is provided for spatializing audio in conference calls, in which the participants in the calls are placed in particular locations in order to provide an additional dimension (direction) so that the participants can better recognize who is speaking. Also, the dimension of volume is provided, which can be used in creating background sub-
10 conferences. Consequently, sub-conferences can be conducted in which the participants can move freely between them and also listen to other sub-conferences being conducted in the background.

 An important technical advantage of the present invention is that by using a spatial layout for a conference call, the audio streams from different sub-
15 conferences are distinguishable, and a user is then able to attend multiple sub-conferences.

 Another important technical advantage of the present invention is that by using a spatial layout for a conference call, one participant will always know which other participant is speaking, as long as the first participant knows the spatial
20 location of the other participant.

 Yet another important technical advantage of the present invention is that the use of a spatial layout for conference calls can be used as an aid for auditory memory. This approach can be useful when a participant is located in a noisy environment and it is difficult to recognize a speaker by voice alone, or one person
25 participates in a conference with other people who are unknown and it is difficult to recognize the speaker by voice alone.

BRIEF DESCRIPTION OF THE DRAWINGS

 A more complete understanding of the method and apparatus of the present invention may be had by reference to the following detailed description when taken

in conjunction with the accompanying drawings wherein:

FIGURE 1A is a diagram that illustrates a centralized conference call system that can be used to provide spatialized audio, in accordance with a preferred embodiment of the present invention;

5 FIGURE 1B is a diagram that illustrates a distributed conference call system that can be used to provide spatialized audio, in accordance with the preferred embodiment of the present invention;

10 FIGURE 2 is a flow diagram of an exemplary method that can be used by a user of a terminal to register with a conference, in accordance with the preferred embodiment of the present invention;

FIGUREs 3A and 3B are related diagrams that illustrate a spatial layout for a conference call, in accordance with the preferred embodiment of the present invention; and

15 FIGUREs 4A and 4B are related diagrams that illustrate a spatial layout for a plurality of sub-conferences, in accordance with the preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE DRAWINGS

20 The preferred embodiment of the present invention and its advantages are best understood by referring to FIGUREs 1-4 of the drawings, like numerals being used for like and corresponding parts of the various drawings.

25 Essentially, in accordance with the present invention, a method for spatializing audio in conference calls is provided in which the participants in the calls are placed in particular locations in order to provide an additional dimension (direction) so that the participants can better recognize who is speaking. Also, the dimension of volume is provided, which can be used in creating background sub-conferences. Consequently, sub-conferences can be conducted in which the participants can move seamlessly between them and also listen to other sub-conferences being conducted in the background.

30 Specifically, as illustrated by the diagrams shown in the related FIGUREs 1A and 1B, there are two types of conference call systems that can provide spatialized

audio: centralized and distributed systems. In a centralized system (e.g., as shown in FIGURE 1A), the spatialization process (i.e., process of giving a direction to the source of the audio stream) takes place at a central location. In a distributed system (e.g., as shown in FIGURE 1B), the spatialization process takes place at each terminal
5 involved in the conference call. Notwithstanding the type of conference call system involved, in accordance with a preferred embodiment of the present invention, each participant in the conference call preferably wears a stereo headset or similar apparatus which is fed by two relatively high-quality audio channels (e.g., > 20 kHz). Consequently, in the preferred embodiment, the conference calls can be implemented
10 with 3-dimensional spatialization.

Referring to the exemplary centralized system 100 shown in FIGURE 1A, a plurality of conference call participants are using terminals (e.g., telephones with headsets) 102, 104, 106, which are connected to a network 108. For example, the exemplary network 108 can be a Public Switched Telephone Network (PSTN), a
15 Public Land Mobile Network (PLMN), or the Internet, and the telephones (102-106) can be fixed telephones, mobile radiotelephones, or Personal Computers (PC's), respectively.

For this embodiment, the system 100 also includes a conference call control unit 110 (e.g., part of a server or similar processing unit in the network) connected to
20 a plurality of command units 112a-n and spatialization units 114a-n. The command units 112 and spatialization units 114 are further connected to a common audio bus 116. More precisely, each command unit 112 can output an audio signal to a conductor in the audio bus 116, and a control (command) signal to the control unit 110. Each spatialization unit 114 can receive a plurality of audio signals from the
25 conductors of the audio bus 116, and output a spatialized audio signal which is coupled to the network 108 and then to the terminals 102-106. As such, the control unit, command units, spatialization units, etc., can be analog or digital units.

In operation, a user of a terminal (102, 104, 106, etc.) can send a voice command or other type of control signal (e.g., DTMF tone) to an assigned command
30 unit (112) via an audio connection. For example, in order for the command unit to distinguish between a voice command and ordinary speech, the user can initially

vocalize a predetermined keyword which is unique and unlikely to be interpreted as ordinary speech. A recognition unit associated with the command unit can recognize the spoken keyword as a voice command. Alternatively, for example, the user can press a button (or key) that sends a unique DTMF code to the command unit. The

5 DTMF code can be recognized by the associated recognition unit as an instruction to interpret subsequent vocalized words as commands. As another example, a user-to-user signalling method of issuing commands can be used. For example, a signalling channel can be used to send a command signal to the command unit. Such signalling channels are available, for example, in ISDN networks and also in mobile

10 communications networks (e.g., in the GSM). As such, in response, the command unit 112 routes the control signal to the control unit 110, and couples any audio signals to a conductor (channel) in the audio bus 116. The control signal instructs the control unit 110 about the user's preferences as to a spatial layout (e.g., formation of a sub-conference, etc.). In response to the user's commands, the control unit 110 sends

15 spatial layout commands to the spatialization units 114, which combine the plurality of audio signals received from the audio bus 116 so as to configure spatial layouts in accordance with the users' preferences. The resulting spatialized audio signals are then coupled to the users' terminals via the network 108.

More specifically, FIGURE 2 is a flow diagram of an exemplary method 150

20 that can be used by a user of a terminal to register with a conference, in accordance with the preferred embodiment of the present invention. At step 152, the user (of a terminal 102, etc.) calls a telephone number associated with the intended conference, and at step 154, the call is routed (via the network 108) to a command unit (e.g., 112a). At step 156, the user enters a dialog with the command unit 112a. During this dialog,

25 the command unit 112a interrogates the user to determine, for example, the user's name, what conference the user intends to be connected to as an active participant, and what conference(s) should remain in the background, etc. At step 158, the control unit 110 connects the user to at least one conductor (channel) of the audio bus 116, and thus determines the relative position of the user in the conference.

30 Referring to the exemplary distributed system 200 shown in FIGURE 1B, a plurality of conference call participants are using terminals (e.g., telephones with

headsets) 202, 204, 206, 208, which are connected to a network 210. Again, for this exemplary embodiment, the network 210 can be a PLMN or the Internet, and the telephones (202-208) can be fixed telephones, mobile radiotelephones or PC's, respectively. Alternatively, the network 210 can be a PSTN, which is technically possible but less likely to be used in reality. For this embodiment, each of the terminals 202-208 can be configured to include a control unit (e.g., 212) and a spatialization unit (e.g., 214). In other words, the audio spatialization in this distributed conference call system 200 is preferably accomplished at the terminals involved in the conference call. Each terminal 202-208 outputs an audio signal (e.g., originating from a microphone for the respective user), which is coupled via the network 210 to a spatialization unit in each of the other terminals involved in the conference call.

In operation (referring to the distributed system shown in FIGURE 1B), with a voice command or other command signal (DTMF tone), a terminal's user (e.g., for terminal 202) inputs a control signal to the control unit (e.g., 212) in that terminal. For example, as described earlier with respect to FIGURE 1A, the user's command can be a unique keyword or DTMF code. Also as described earlier, a user-to-user signalling method of issuing commands can be used (e.g., using a signalling channel to convey a user's commands). The resulting control signal instructs the control unit 212 about the user's preferences as to a spatial layout (e.g., create a sub-conference, etc.). In response to the user's command, the control unit 212 sends a spatial layout command signal to the spatialization unit 214, which combines the plurality of audio signals received from the other terminals involved in the call, so as to configure a spatial layout in accordance with the user's preference. The resulting spatialized audio signals are then coupled to the user's (stereo) headset.

In accordance with the present invention, a user can identify a plurality of sub-conferences by their relative spatial locations. The user can then select one of those conferences for active participation. For example, a list of identifiers associated with the sub-conferences and information about their relative positions can be displayed by a terminal (e.g., using a PC via the Internet). The user then selects from the list in order to participate in one or more of the sub-conferences. Alternatively, for example,

the user can initiate a procedure that browses the audio channels, whereby the system couples the user to each of a succession of different sub-conferences (e.g., brings each of the sub-conferences to the foreground in succession). As such, for example, during a certain period of time, the user can select one (or more) of those sub-conferences for active participation (e.g., by sending a DTMF signal or voice command to the control unit 110 or 212).

FIGURES 3A and 3B are related diagrams that illustrate a spatial layout for a conference call, in accordance with the preferred embodiment of the present invention. Essentially, a round table approach can be used for selecting the position of the participants involved in the conference call. In other words, for this embodiment, there is a consistent left-to-right order used for the participants involved. As such, referring to FIGURE 3A, from a first participant's (A) point of view with respect to the spatial layout of the conference, that participant (A) is located in the center of a circle, and the other participants (B-E) are located in the half circle in "front" of the first participant (A). Similarly, referring to FIGURE 3B, from participant E's point of view (maintaining the left-to-right order), that participant (E) is located in the center of the circle, while the other participants (A-D) are in "front" of participant E. Such a layout is preferable, because people prefer to hold conversations with other people who are in front of them rather than behind them. Nevertheless, although a left-to-right order is used for the preferred embodiment, the invention is not intended to be so limited, and in a different embodiment, a right-to-left order can be used.

FIGURES 4A and 4B are related diagrams that illustrate a spatial layout for a plurality of sub-conferences, in accordance with the preferred embodiment of the present invention. Referring to FIGURE 4A, a plurality of sub-conferences 302, 304, 306, 308, 310 are shown for an exemplary spatial layout. The sub-conferences can be created by a user sending a command signal (e.g., for voice recognition or DTMF tone) to the appropriate control unit (110 or 212) shown in FIGURE 1A or 1B. If a user desires to leave a current sub-conference, the user can send a "leave" command to the appropriate control unit. That user (e.g., user x in FIGURE 4A) is moved to the outer circle where all other sub-conferences are located in the spatial layout. As shown, the user x is not participating in any conference or sub-conference, and users

A-E are participating in the same sub-conference 302.

In order to participate in another sub-conference, a user can send an "approach <argument>" command to the appropriate control unit. The "argument" is the identifier of one of the participants in that other sub-conference. The user will be placed in "front" of the participant identified in the "argument" in the spatial layout of that sub-conference, and then the user can begin to participate in that sub-conference. Notably, the user can also participate in another sub-conference without having to send a "leave" command to the control unit. Note that all of the sub-conferences are located at a relatively large spatial distance to any other sub-conference. Consequently, a participant in one sub-conference will hear all other participants in that sub-conference in the foreground, and the participants in other sub-conferences in the background. As illustrated by FIGURE 4B, if a user (A) does not participate in any sub-conference, that user can be spatially located as shown. As such, that user (A) can listen to all of the sub-conferences (304-312) simultaneously, and can join and participate in one of the sub-conferences as desired.

In accordance with the present invention, by using such sub-conferences in a spatial layout as shown, there is less need to appoint a conference moderator (nevertheless, moderators will still likely be used for the sub-conferences). In any event, the participants can create sub-conferences themselves if so desired, and move between the sub-conferences freely. As such, using the present invention's spatial layout approach, the conference/sub-conference participants can be in effect their own moderators. Furthermore, by using such spatial layouts, users can attend multiple conferences as long as the audio stream from the conferences are distinguishable. In this regard, human beings are capable of monitoring numerous conversations simultaneously, and can focus on any one of the conversations while placing the other conversations in the background. This cognitive phenomenon, which the present invention preferably takes advantage of, is the so-called "Cocktail Party Effect". In other words, using a stereo headset along with the present invention's spatial layout approach, for example, a conference/sub-conference participant can distinguish audio streams from other participants due to their different spatial locations. The audio streams can appear to be coming from different directions (e.g., from different

locations in a sub-conference), or to be originating at different distances (e.g., from other sub-conferences).

5 In order to create stereo sound in the preferred embodiment, the system can use pairs of audio channels. For example, in a fixed public network, an ISDN connection can provide two channels that can be used to provide stereo sound. Similarly, for example, a wideband cellular system can assign two channels for each such stereo connection to be used.

10 Although a preferred embodiment of the method and apparatus of the present invention has been illustrated in the accompanying Drawings and described in the foregoing Detailed Description, it will be understood that the invention is not limited to the embodiment disclosed, but is capable of numerous rearrangements, modifications and substitutions without departing from the spirit of the invention as set forth and defined by the following claims.

-10-

WHAT IS CLAIMED IS:

1. A system for distinguishing between participants in a conference call, comprising:

5 a plurality of terminals, each one of said plurality of terminals configured to send an audio signal and receive a spatial audio signal, said plurality of audio signals creating said conference call;

a control unit coupled to said plurality of terminals, said control unit configured to output a spatial layout signal responsive to at least one command signal from said plurality of terminals; and

10 a plurality of audio spatialization units, each one of said plurality of audio spatialization units coupled to said control unit and said plurality of terminals, at least one of said plurality of audio spatialization units configured to output said spatial audio signal responsive to said spatial layout signal.

2. The system of Claim 1, wherein said at least one command signal
15 includes a voice command associated with said spatial layout signal.

3. The system of Claim 1, wherein said at least one command signal includes a DTMF tone associated with said spatial layout signal.

4. The system of Claim 2 or 3, wherein said at least one command signal includes a data signal transmitted over a signalling channel.

20 5. The system of Claim 1, further comprising a command recognition unit coupled to at least one of said plurality of terminals and said control unit.

6. The system of Claim 1, wherein each one of said plurality of terminals is coupled to said control unit and said at least one of said plurality of audio spatialization units by a telecommunications network.

25 7. The system of Claim 6, wherein said telecommunications network

-11-

comprises a PSTN.

8. The system of Claim 6, wherein said telecommunications network comprises a PLMN.

5 9. The system of Claim 6, wherein said telecommunications network comprises an Internet.

10. The system of Claim 5, wherein said command recognition unit includes a voice recognition circuit.

11. The system of Claim 5, wherein said command recognition unit includes a tone recognition circuit.

10 12. The system of Claim 5, wherein said command recognition unit includes data signalling receiving circuitry.

13. The system of Claim 1, wherein said spatial layout signal determines a position for each of said participants in said conference call.

15 14. The system of Claim 1, wherein said spatial layout signal determines a position for at least one of said participants outside of said conference call.

15. The system of Claim 1, wherein said spatial layout signal comprises a left-to-right order for said participants in said conference call.

20 16. The system of Claim 1, wherein said plurality of terminals, said control unit, and said plurality of audio spatialization units comprise analog terminals and units.

17. The system of Claim 1, wherein said plurality of terminals, said control

-12-

unit, and said plurality of audio spatialization units comprise digital terminals and units.

18. A terminal for use in distinguishing between participants in a conference call, comprising:

5 a control unit configured to output a spatial layout signal responsive to a first participant's preference; and

an audio spatialization unit coupled to said control unit, said audio spatialization unit configured to receive a plurality of audio signals from terminals associated with other participants, and output a spatial audio signal comprising said
10 plurality of audio signals arranged in response to said spatial layout signal.

19. The terminal of Claim 18, wherein said audio spatialization unit is coupled to said terminals associated with said other participants by a telecommunications network.

15 20. The terminal of Claim 19, wherein said telecommunications network comprises a PSTN.

21. The terminal of Claim 19, wherein said telecommunications network comprises a PLMN.

20 22. The terminal of Claim 19, wherein said telecommunications network comprises an Internet.

23. The terminal of Claim 18, further comprising an analog terminal.

24. The terminal of Claim 18, further comprising a digital terminal.

25. A method for distinguishing between participants in a conference call, comprising:

25 each one of a plurality of terminals outputting an audio signal, said plurality

-13-

of audio signals creating said conference call;

generating a spatial layout signal responsive to at least one command signal;

and

generating a spatial audio signal responsive to said spatial layout signal, said

5 spatial audio signal comprising said plurality of audio signals.

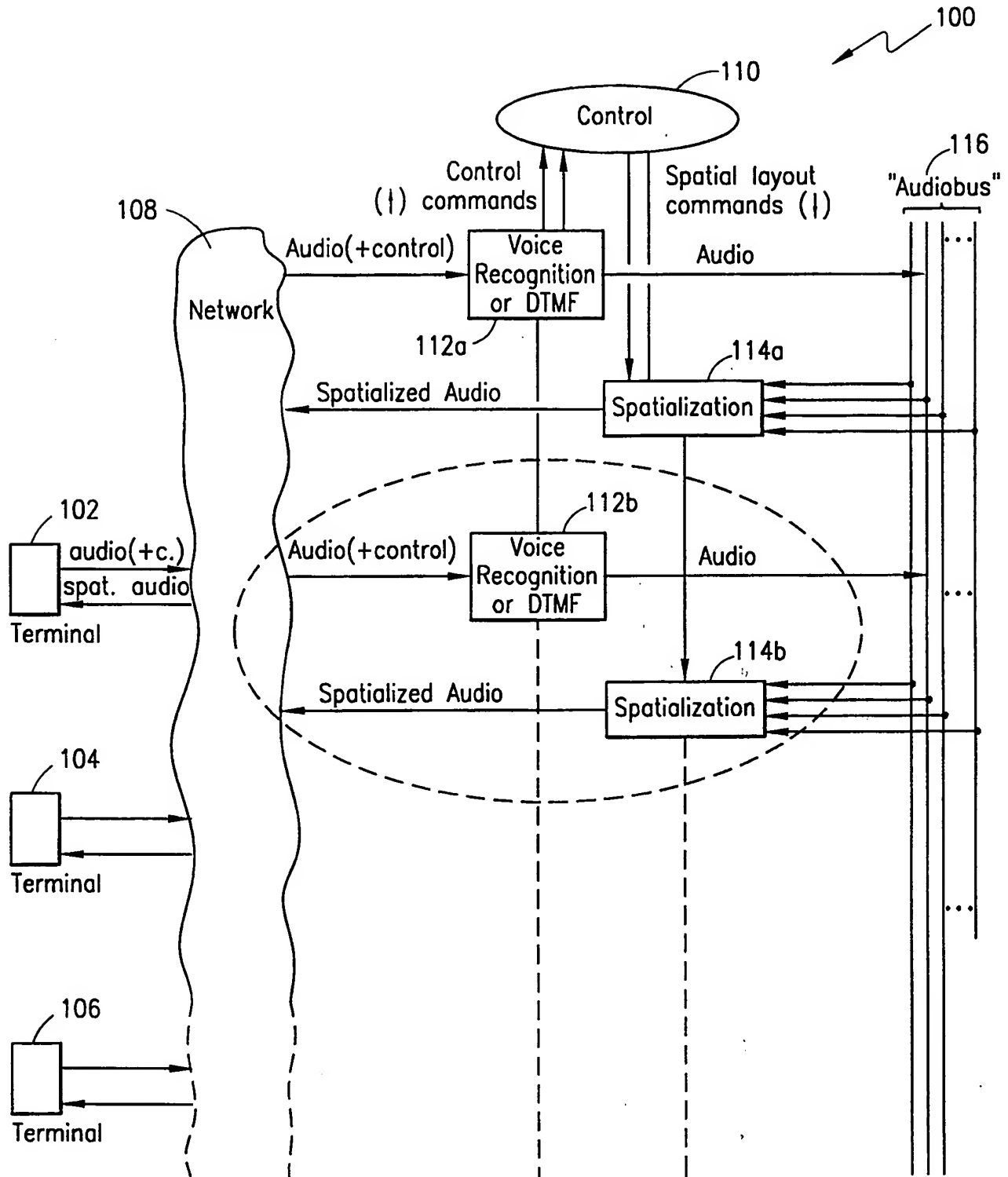
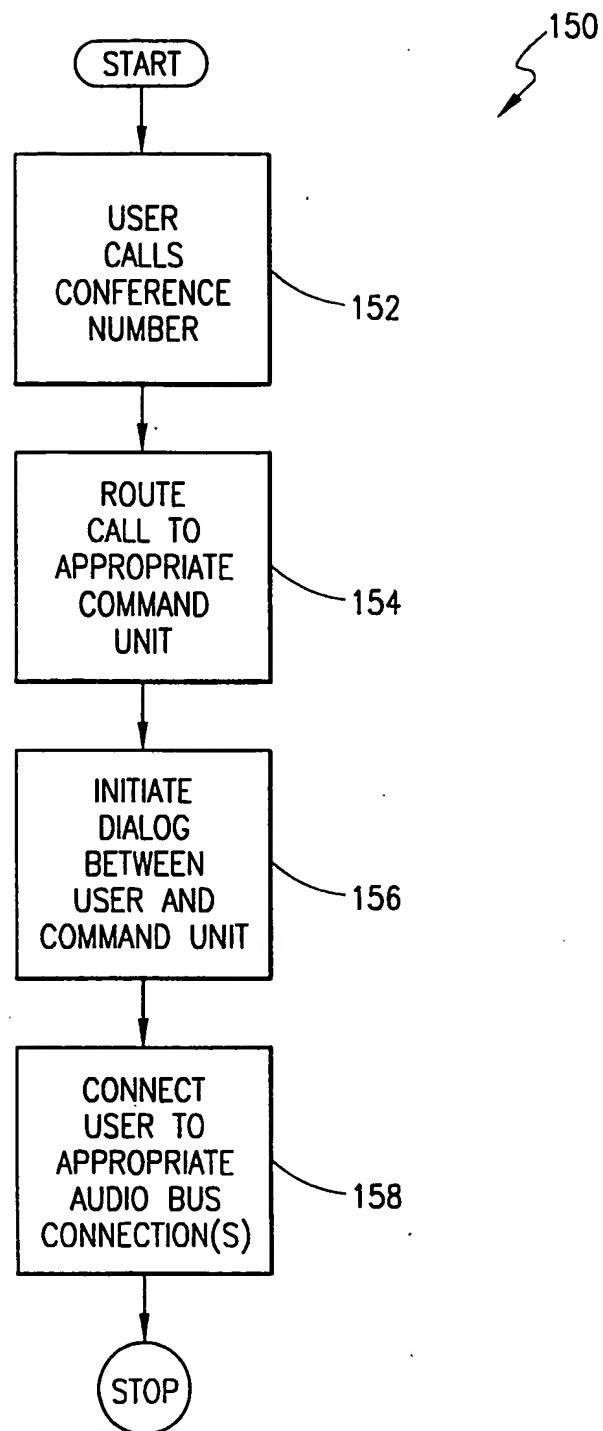
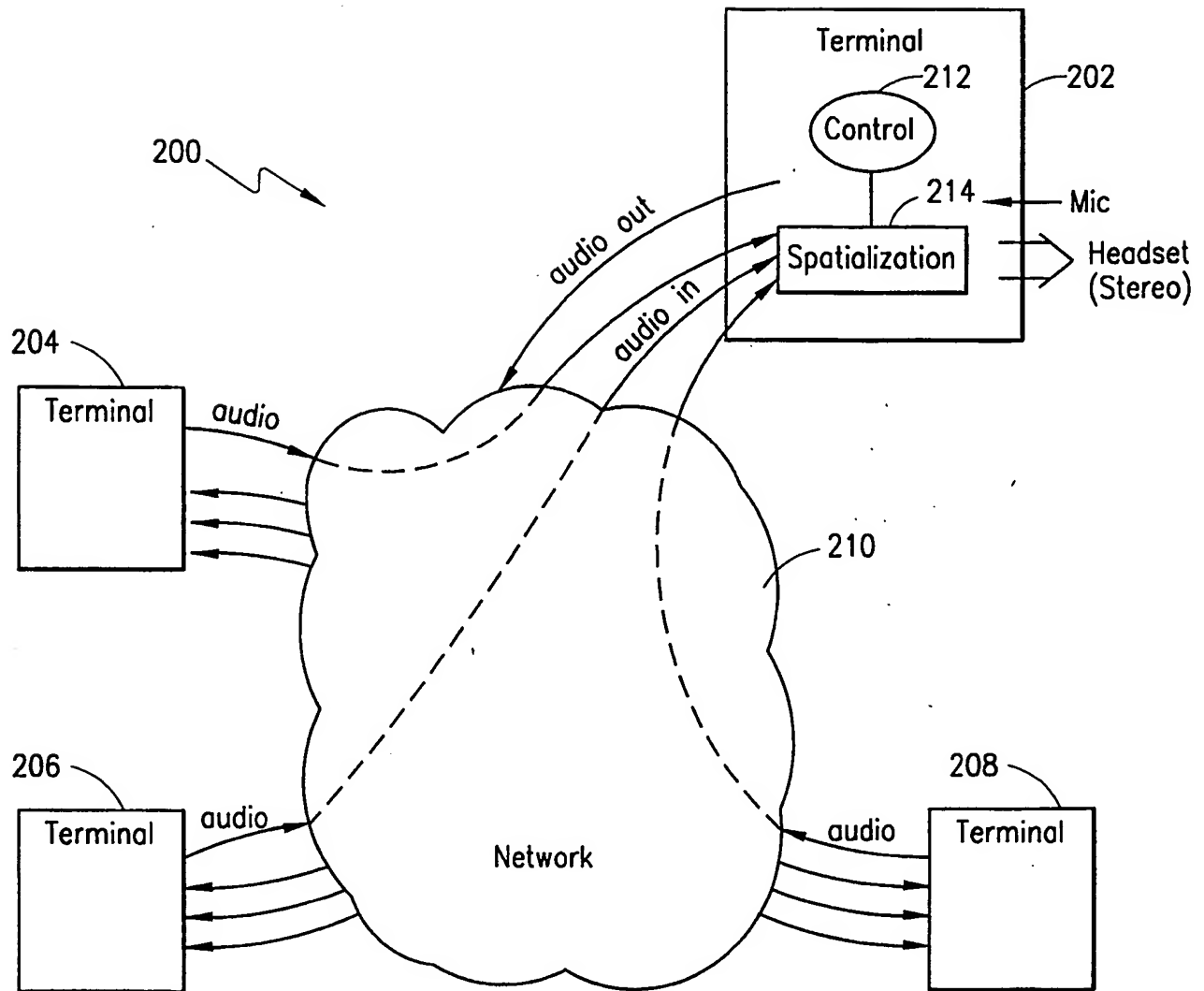
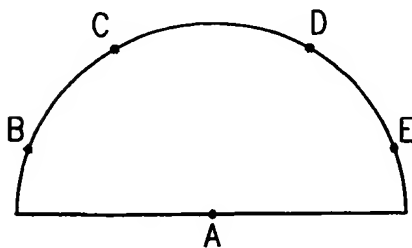
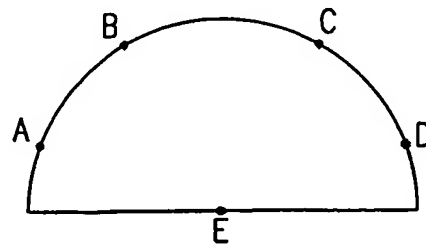


FIG. 1A

**FIG. 2**

*FIG. 1B**FIG. 3A**FIG. 3B*

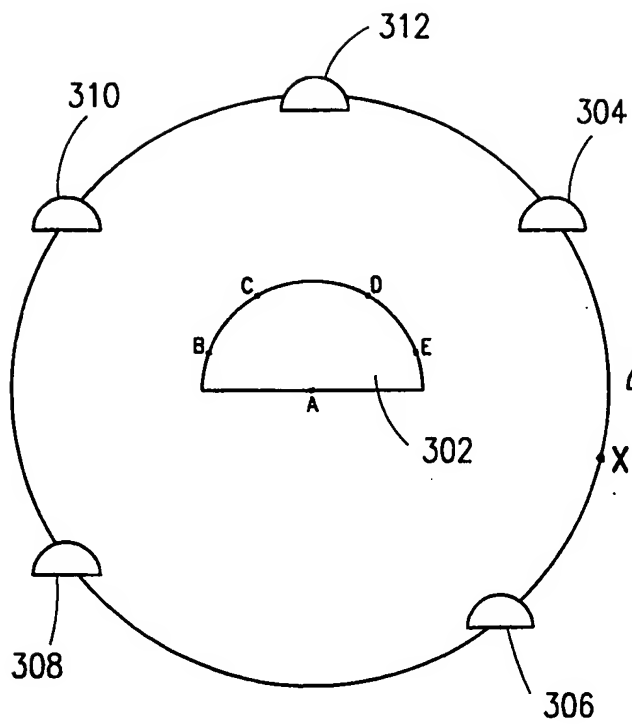


FIG. 4A

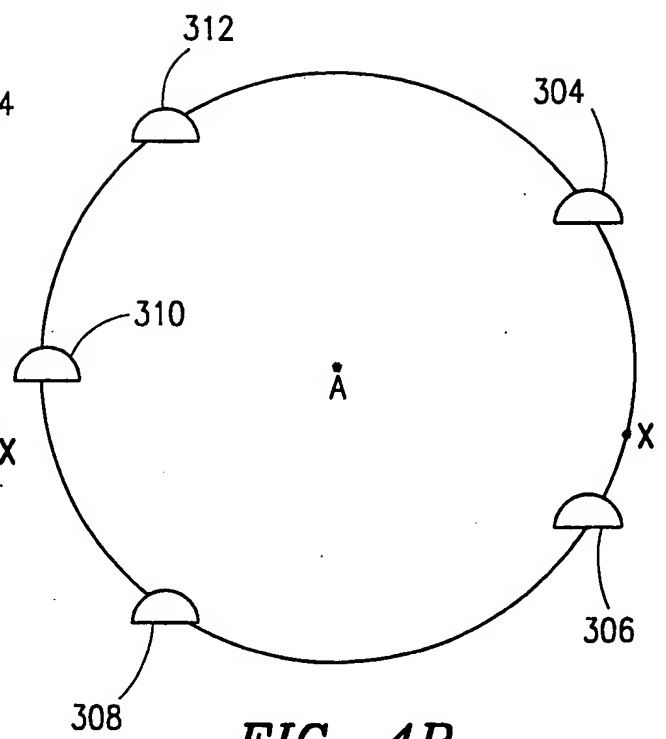


FIG. 4B

INTERNATIONAL SEARCH REPORT

International Application No

PC 00/00204

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04M3/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GB 2 303 516 A (PLESSEY TELECOMM) 19 February 1997 (1997-02-19)	1, 6, 7, 13-15, 17-20, 24
A	page 1, line 15 - line 19 page 2, line 17 - line 23 page 3, line 13 - line 16 page 10, line 7 - line 15 page 11, line 6 - page 12, line 3 figure 4	3-5, 11, 12
A	--- US 5 020 098 A (CELLI JOHN R) 28 May 1991 (1991-05-28) column 1, line 32 - line 38 column 1, line 31 - column 2, line 17 column 3, line 1 - line 25 --- -/-	1, 6-9, 13, 17-22, 24

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

14 June 2000

Date of mailing of the international search report

20/06/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Vaucois, X

INTERNATIONAL SEARCH REPORT

Int. Patent Application No

PC 00/00204

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4 734 934 A (BOGGS GEORGE J ET AL) 29 March 1988 (1988-03-29) column 1, line 5 - line 50 column 4, line 48 - line 55 ---	1,6,7, 13,16, 18-20,23
A	RADAMIS BOTROS, ONSY ABDEL-ALIM, PETER DAMASKE: "Stereophonic speech teleconferencing" ICASSP 86 PROCEEDINGS. IEEE-IECEJ-ASJ INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (CAT. NO.86CH2243-4), 7 - 11 April 1986, pages 26.6.1-26.6.4, XP002107517 page 26.6.1, left-hand column, line 11 - line 19 page 26.6.1, right-hand column, line 24 - line 36 figure 2 -----	1,6,7, 16, 18-20, 23,25

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No

PCT/JP 00/00204

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
GB 2303516	A	19-02-1997	EP 0762717 A	12-03-1997
US 5020098	A	28-05-1991	CA 2027212 A,C	04-05-1991
US 4734934	A	29-03-1988	NONE	